The 18th International Conference on Advanced Information Networking and Applications (AINA 2004) March 29 - March 31, 2004, Fukuoka Institute of Technology (FIT), Fukuoka, Japan, to appear in IEEE Computer Society proceedings.

Mixer-Subverter :: an Online Improvisational Video System

Paul Nemirovsky
MIT Media Lab
pauln@media.mit.edu

Cati Vaucelle
Media Lab Europe
cati@mle.ie

Ariadna Quattoni
MIT AI Lab
ariadna@ai.mit.edu

Abstract

This paper describes the Mixer-Subverter; an online system that allows children to integrate the activities of play (from giving to stealing; from sharing to forcing to receive) and the activities of video editing (creating, juxtaposing, controlling) into a never-ending process of mix and subversion. It invites the storyteller within each one of us to compose and visualize movies, images and sound environments while writing a story. In addition, the Mixer-Subverter encourages playful collaboration in an exchange network of unique media artifacts.

The operation of the system is based on improvisational principles; an idea that there is not a particular plan or goal to the editing process. Instead, pieces that populate the Mixer-Subverter's media space acquire their meaning through their patterns of use and practices of exchange [1, 2]. This paper is a report on a work in progress. As such, it presents the underlying rationale and provides a description of the first prototype version of the system.

1. Scenario

Three children are playing in a park next to a small pond. With them there is an ensemble of objects of different colors and shapes: cars and ships, stones and sticks and leaves, all scattered around the pond. The children pick a few of the objects and start playing. As they go on, making stories along the way, each of them gets excited about a different set of actions: building or navigating around dams, hideouts, and bridges, or destroying the dams and bridges of others. Being completely engaged in the play, they nevertheless have only a vague and continuously changing notion of "what the play is about"; each childís story is his own [3]. Yet, something brings them to play together, to weave stories where they have not a goal and not a plan for how the story will develop. Why do they feel like engaging in such ad-hoc collective storytelling? On the first glance, their ways of playing with each other are hardly what we would describe as "collective" or "shared" development. It has none of that pacifying yet boring spirit of predictability in the interaction; they steal game artifacts from each other, they fight over the territory, they consider every element of the game as unique and they are, for the most of the time, unwilling to let it go. Frequently, they try to impose a story of their making on others, yet, not willing to play alone, they choose to participate in stories that are only partially their own. Such collective story-making comes about through exchanges of objects and plots, an exchange that can be seen as improvisational in its nature, for no well-defined plan or goal is pursued and the focus is on the process of exploring the story, rather than creating a polished end story artifact. The type of story that emerges as a result is rarely the domesticated creature they encounter in school, the kind of knowledge exploration made boring by its linearity and predictability. Their game in the park does not consist of sitting together and making up one nicely arranged story; rather, the story is a part of the game itself, a continuously evolving process ongoing in the childrenís minds.

How can we bring this experience of unplanned play into the digital domain, and more precisely, into the process of video creation and exchange? To do that, we need to engage our audience in a never-ending process of story-making. Unlike a compositional story-making, one that focuses on simply combining story particles in a particular and fixed order, such improvisational story-making needs to be able to focus on a continuous change in the connections and properties of the story particles and their exchange between the participants.

... Juan, 15 years old, has just shot his latest soccer match. Having digitized the entire movie, he connects to the Mixer-Subverter; a system that allows him to mix his own and othersí video clips in a subversive manner. While his movie is being played back, and

with the system in the *learning* mode, he describes the video with sentences such as "I was so close to the goal… then I got the ball and ran… They could not resist that shot!" The system learns the association of the description and dynamically creates cue points within the movie, segmenting it into relevant 'virtual' clips. These clips form Juan's palette of available movies. Later, having switched to the *subverting* mode, he will type "close to the goal" to bring forward that cool shot – or a shot that has been placed in its place by other participants of the Mixer-Subverter. These other participants can also use Juan's clip; in fact, the more Juan's clip is used in the movies made within the system, the more valuable it becomes. As its value increases, more participants might become interested in getting their hands on it. Isabelle, 13 years old, will try to exchange her best shot about her dog falling down from a chair, to get Juan's great soccer shot. If Juan wants to keep the shot, he will have to protect it whenever a special type of a 'virus' within the Mixer-Subverter's space creates an "open door" opportunity for everyone to get anything they want from Juan's palette of video clips.

In this scenario story-making becomes a game in which each element within the system gains value by its context of use. Given a foundational rule of the Mixer-Subverter, which states that only a single instance of every object can exist within the system at any given point, that object's history of exchange becomes a key element of the game. The basic principle of exchange, as described by Mauss [1], is incorporated into the Mixer-Subverter: by giving an object away (in our case, a video segment), one's ownership rights are extended to include the artifacts owned by the receiving side (exemplified, in this case, by a right to build connections to the other participant's collection). Such mechanism of exchange brings dynamicity to the storytelling process, connecting the images to the collective imagination and making explicit the value afforded to these images by the community.

2. Introduction

definition: Mixer-Subverter (hereafter M-S) is a command-driven, responsive system which mediates video segments and their interconnections in real time, between a community of net-based participants.

purpose: M-S aims to provide an environment wherein children can create and share networks of unique video artifacts. The purpose behind creating such an environment is to support the type of collective storytelling that makes use of children's natural improvisational tendencies (or intelligence, in Piaget's [4] terms). Using the M-S, children can improvise movies in real time, both privately and collectively. These movies are automatically segmented and associated with keywords, based on the content of mini-stories provided by the participants. Subsequently, the clips are *likely* to be recalled by typing the same stories, with the likelihood of recall dependent on the amount of subversion taking place within the M-S network.

motivation: M-S aims to bring improvisational spontaneity into the process of movie creation; a notion that runs contrary to the parametric video editors of today. It serves as our platform for creating a video-oriented interaction along the lines of the play described in the Scenario, by encouraging associative thinking; one where media exchange leads the children to a different type of "day dream" [5]. To quote Bachelard, "…imagination is always considered to be the faculty of forming images, but it is rather the faculty of deforming the images, of freeing ourselves from the immediate images; it is especially the faculty of changing images." [6]

contribution: M-S invites interaction between the audience and the story rather than lots of preplanning which usually characterizes the delivery of a controlled message. M-S allows variations in the story form, based on the narrative imagination of the participant. Pictures may lead the fantasy of the participant during this improvisation; the dynamic of the video shown might stimulate an influx of memories. It is up to the participants to dynamically improvise any association of words and images – and later, with the computer now recognizing this particular association, collaborate and compete with other participants in creating movies coming as a result of stories being made on the fly.

3. Past and Present Solutions

Most immediately, Mixer-Subverter is a continuation of the authors' previous work focusing on improvisational methods for media creation, browsing, annotation, and performance. Among the directly relevant projects are the Emonic Environment [7], a system that allows its users to create, navigate, and evolve media networks in an improvisational fashion; Textable Movie [8] that allows its users to annotate images in order to retrieve them later within a specific context; and adaptive systems such as [9] that extend the concept to the making of movies with a number of

automatic functions. All these works, while quite different in their approach and application, have a similar concern in mind: video editing tools available today provide its users with many parametric functions for creating movies, but do so while placing rigid constraints on spontaneity.

More generally, we are inspired by the work of Nam Jun Paik, Woody Vasulka, Christa Sommerer, and many others who, each in their own way, broke with the norms of traditional video editing, taking a conceptually different approach; one where the objective is engaging in a process, rather than creating an artifact.

4. Challenges

Faced with the task of developing a model for creation and exchange of video in an improvisational manner, we have to consider a large number of issues having to do with the nature of improvisation and improvisational participation, the context of the medium and the social setting to which our model would apply. As a result, the following list is by no means complete; instead, it highlights what we consider to be the most relevant points of the improvisational model being created:

1st: Dynamic nature of improvisational structures. Structural representations used in the course of improvisation are to be incorporated, modified, and purged dynamically to satisfy the improviser's changing goals and attention. As the criteria guiding improviser's behavior (in terms of expectation and evaluation procedures he employs) evolve in the course of a performance, he changes what he considers the "right thing to do". This change comes based on the combination of these evolving criteria and the stimuli from the environment, manifesting themselves only as the performance develops.

2nd: Changing, multileveled focus. An improviser thinks about what he's doing at different levels of abstraction simultaneously. Continuously switching between macro- and micro-level, he attends to the very minute (e.g. a particular RGB mask) at one moment, only to switch and think about structural development (e.g. a climax) a second later.

3rd: Diversity of types. Improvisation is a result of interrelating multiple perceptual inputs and memories; an improvisation whose 'output' is video is nevertheless an improvisation that includes auditory, tactile, and other formative content. An image might be inspired by a sound, which in turn is inspired by a text or another image; this free and proactive interaction of types is integral to the improvisational process.

4th: Relevance of context. Following on the above point, the improviser's decision-making is rooted in the totality of his perception of the moment. Thus medium-specific laws of decision-making should be used cautiously in deciding the subsequent output, for the perception of any media is in itself an act shaped by the context. Indeed, improvisation is not formed in a vacuum or in one medium separated from others; it strives to incorporate or reflect the environment in which it is created.

5th: Process, not artifact production, as the goal. An improviser, unlike a feature-film cinematographer, a Western composer or a product designer, is not concerned with the production of a final artifact – a movie, a sonata, a pop-song, or a chair. While improvisation might be recorded and, as such, seen as a fixed construct, the true point of improvisation is the process of exploration, contextualizing and interrelating memories and perceptions (e.g. children's play in the park, mentioned in the Scenario, has as its objective the play itself, rather than a story as a finalized product). An improviser's job is to weave together an array of 'sketches', which gain their relevance (and meaning) only as the improvisation unfolds.

6th: Absence of a static plan. Planning, in its traditional sense, does not seem to be the optimal way to think about the process of improvisational creation. Instead, the act of improvisation might be more aptly described as one of exploration and continuous evolution of multiple mini-plans. In other words, an improviser is usually far less concerned with perfectly playing to a specification than he is with breaking a new ground and learning from unintended mistakes and unexpected successes.

7th: Issues of control and responsibility. In an improvisational performance, no fixed contract specifying responsibilities of control (balance of power) exists between the performers; the criteria that define the degree to which each party assumes creative control over different aspects of the ongoing improvisation are set dynamically, according to both implicit and explicit negotiations between the performers. Giving up part of the control also frees the improviser from the preoccupation with creating a perfect finalized product. In other words,

improvisation implies a lower cost of experimentation, allowing spontaneous exploration of new, 'unproven', ideas.

8th: Continual feedback. Improvisation is not evaluated at one point in time or space. Over the course of a performance, improvisers provide feedback to each other. This feedback ranges from general and vague to particular and precise; what defines its value is the ability of the recipients to learn from it and move in new directions. The learning is not procedural; it cannot be summarized by a symbolic rule. Instead, it can be described as discovering patterns where one didn't see them before.

9th: Meaning-making through exchange. In an improvisational group action, construction of meaning happens through the exchange of elements. In other words, an image or a sound acquires its meaning only through the details of its history of use – where and how it has been employed before. These details determine how it or similar elements are perceived the next time they are encountered.

10th: Audience as a participant. From the passive audience of the linear storytelling to the nearly equally passive audience of the multiple-choice "interactive" environments, a strict giver / taker dichotomy has been enforced between the consumer (the audience) and the producer (the performer). In an improvisational network, however, such a distinction is obsolete; anyone can co-improvise, so long as the effect of his activity is seen / heard in one way or another by the other performers. Similarly, even when not actively participating in the act of media creation, the audience is not to be regarded as passive; it is to be viewed as a part of the improvisational circle.

11th: Ownership. Ownership is an essential part of a truly engaging storytelling – for one of its important aspects is the recreation of the self in a fictional context. Ownership thus must be retained even in the improvisational context for the participants to feel that they own some part of the story, its values. It has however to be coupled with the idea of exchange (i.e. extending one's ownership) to reflect the fluid nature of ownership in the course of improvisation.

5. Implementation

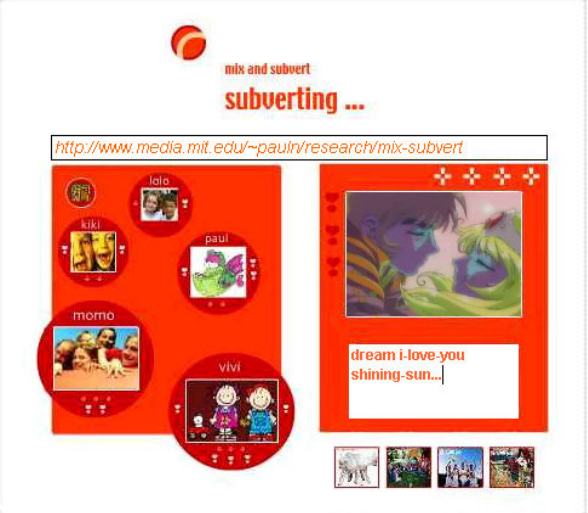

Figure 1. Mixer-Subverter prototype UI

palette: The first prototype of the Mixer-Subverter has been implemented in Lingo; it runs as a Macromedia Shockwave applet and is accessible on the Internet. This prototype version is capable of indexing and segmenting an unlimited number of movies based on the descriptions provided by M-S's participants in real time, thus allowing each participant to populate his 'own' video palette. Once a palette has been populated with movies and descriptive sentences, the participant can create his own movies. Word recognition is dynamic and simple pattern matching can be performed on the fly. Some functional commands can also be used (e.g. RGB modifiers, scale, and speed effects) in controlling the video stream. The system currently has three modes:

- *teaching* mode: segmenting a big movie into chunks, created dynamically by observing the participant's keyboard input.
- *action-provoker* mode: text snippets in the spirit of the Happenstance movement (Nyman, 1976), sent to the participants by the system, suggesting possible courses of action in acquiring new footage.
- *subverting* mode: using the palette, stored words, and sentences to assemble a new movie on the fly.

The palette is invisible to other participants; only the movie that is the outcome of the palette's use. In this way, the definition of a single segment as specified by the owner might differ from its perception by other participants, who see it only in the resulting context. This discrepancy allows for continuous evolution of the 'core' video segments.

uniqueness of elements: video segments within the M-S are always unique; they have no copies. Similar to other games (but unlike other media environments), the system has only a limited number of artifacts (ex. cards) of a given type at any given time (as to not destroy the notion of a game). Such a lack of elements keeps up the desire to get hold of the popular ones.

extending the modes: the current basic prototype functionality is being expanded to include the following types of action: fight over a segment, steal from a broadcast, break into a broadcast, connect to another participant's palette. Due to the uniqueness requirement described above, all the operations that involve getting a video segment from another participant also implicitly involve the deletion of that segment from its original location. Additionally, a computer-driven method for subversion of the participants' video is also being developed: 'viral video artifacts', snippets of video used by the M-S itself to algorithmically interject and subvert participant-made stories.

rating the segments & connections: The M-S is aimed at encouraging children creativity rather than merely facilitating the construction of video palettes. To achieve this aim, each video segment has two types of rating: personal and social. Both ratings are influenced by the segment's pattern of use. If a participant uses a video segment, its personal rating goes up. Additionally, if an element is exchanged a lot between the participants, its social rating increases. Rewarding for an increased rate of segments' exchange is aimed at motivating dynamism within the network, encouraging collective creativity.

The ratings are visualized as virtual watermarks shown on top of any footage. The visualization is a three-part shape consisting of the M-S symbol and two stars. The stars represent the two axes along which each video segment is evaluated: the personal and the social. The brightness of the stars reflects the timeliness of the rating – the brighter the stars are, the more recent the rating is.

mystery door: The *door* is the M-S's access point to a given participant's palette. Normally the door remains locked. To encourage the exchange of video segments and their interconnections, every few minutes (a randomly varying interval), M-S issues an alert: "In one minute, someone's door will open up". Nobody knows whose door that is. During the time the door is opened (designed to be less than a minute) everyone might try and steal from the participant with the unlocked door. At the same time however, this participant also has the power to steal anything he wants from anyone else. The idea behind the mystery door is simple: traditionally, transparency is valued in the UI design. With its 'everything is naked, nothing is hidden' agenda, it is however frequently boring and predictable; not the best recipe for creative work. By adding the door we hope to bring a bit of mystery into the M-S's creative process.

6. Conclusions

In this paper we argued that there is a need for an alternative framework for video editing and sharing. The need is motivated by the desire to reproduce the playful improvisational environment of child storytelling in a digital network. We have outlined the rationale for the creation of the Mixer-Subverter, our implementation of such a framework, and described its current initial prototype functionality.

Our plans include a further analysis of the new possibilities for creativity that such a framework can offer as well as evaluating the reception of M-S by participants of different ages, but especially children. To this extent, our immediate future plans include conducting a user study with a group of 8-14 year old children playing with the prototype system. The study (which is currently being designed) will have both quantitative and interview-form parts.

7. References

[1] Mauss, M. (1954). *The Gift: The Form and Reason for Exchange in Archaic Societies*. Glencoe, Ill., Free Press.

[2] Malinovsky, B. (1922). *Argonauts of the western Pacific.* London, G. Routledge & Sons, ltd.

[3] Ryokai, K., Cassell, J. (1999). "StoryMat: A Play Space with Narrative Memories". In Proceedings of IUI '99, ACM.

[4] Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: International Universities Press.

[5, 6] Bachelard, G. (1971) *On Poetic Imagination and Reverie.* Indianapolis: Bobbs Merrill.

[7] Nemirovsky, P., Watson, R. "Genetic Improvisation Model: a framework for real-time performance environments". In the proceedings of EvoGP 2003.

[8] Vaucelle, C., "Textable Movie: improvising with a personal movie database". SIGGRAPH 2003, Conference Abstracts and Applications.

[9] Lieberman H., Liu H. (2002). "Adaptive Linking between Text and Photos Using Common Sense Reasoning". CAH-AWS.